Loan Risk Analysis Project – Technical Documentation

1. Project Overview

This project focuses on analyzing customer loan data to identify patterns that lead to loan defaults. The goal is to support financial institutions in making data-driven and risk-aware lending decisions.

2. Business Problem

Loan defaults cause significant financial loss to banks. By analyzing historical loan data, we aim to identify high-risk customer segments before loan approval.

3. Dataset Description

The dataset contains customer demographic, financial, and loan-related attributes such as income, loan amount, credit score, employment type, and loan status.

4. Tools & Technologies

SQL (MySQL) was used for querying and cleaning data. Python (Pandas, NumPy, Matplotlib) was used for EDA. Power BI was used for creating interactive dashboards.

5. Data Cleaning Steps

Missing values were handled, duplicate records were removed, invalid values were corrected, and data types were standardized to ensure accurate analysis.

6. Exploratory Data Analysis (EDA)

EDA was performed to understand distributions, correlations, and relationships between customer attributes and loan default behavior.

7. Key Insights

Customers with low credit scores and high loan amounts showed higher default probability. Self-employed applicants were identified as a higher-risk group.

8. Business Recommendations

Banks should apply stricter credit checks for high-risk profiles, introduce risk-based interest rates, and closely monitor self-employed applicants.

9. Conclusion

This project demonstrates how data analysis can reduce financial risk and improve decision-making in the lending process.